

LTDP project at INFN CNAF

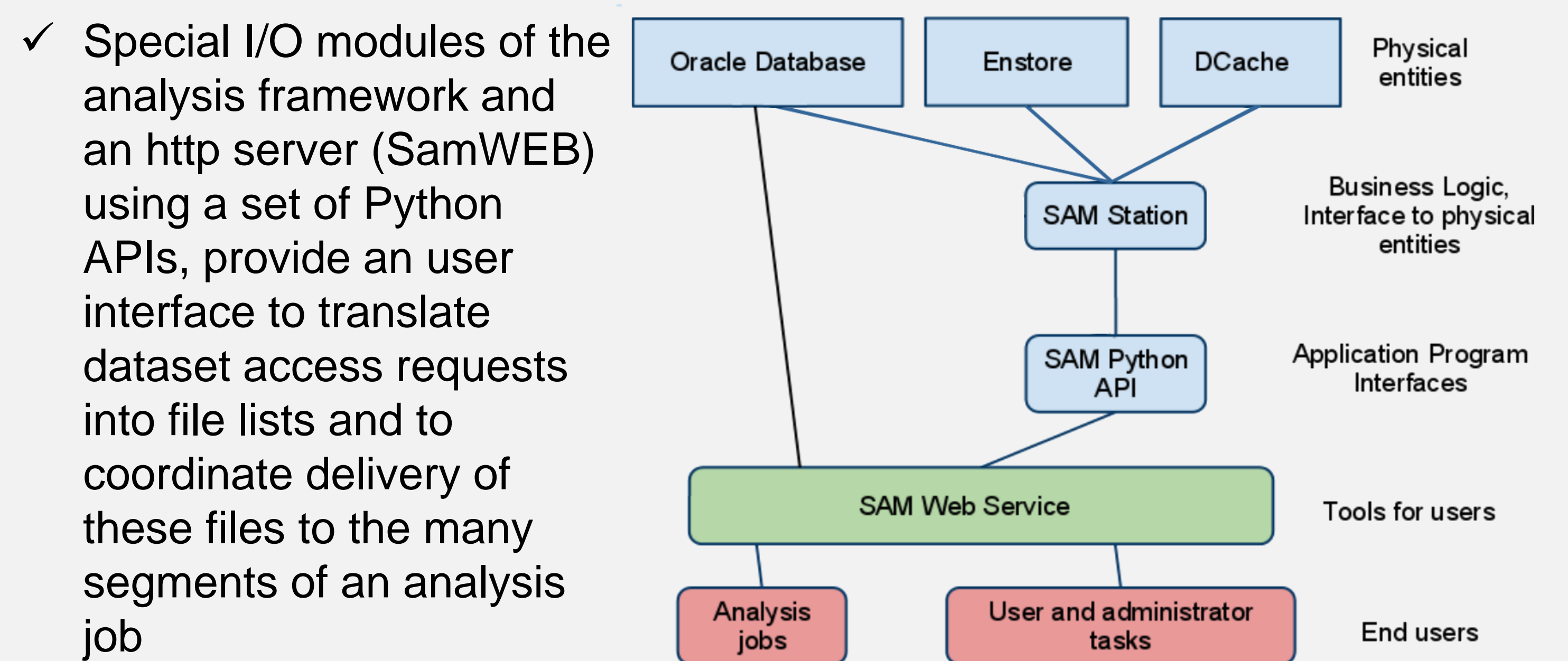
During the latest years, **INFN CNAF** has been working on the **Long Term Data Preservation (LTDP)** project for the **CDF** (Collider Detector at Fermilab) experiment, active at **FNAL** (Fermi National Accelerator Laboratory) from 1990 to 2011

The main aims of the project are:

- ✓ **bit preservation** - to protect data of the CDF RUN-2 data collected between 2001 and 2011 and already stored on CNAF tapes
 - ✓ 4 PB of raw data and analysis-level n-tuples (the most relevant datasets of RUN-2)
 - ✓ periodic check of data availability and consistency
- ✓ **framework preservation** - to ensure the availability and the access to the analysis facility of those data over time
 - ✓ implying the possibility for previously authorized scientific communities to access and use data through dedicated software services

CDF data handling at FNAL

- ✓ At FNAL, data are stored on tape managed by Enstore and accessed through dCache
- ✓ The Serial Access to Meta-data (SAM) station is a set of code products which interacts with an Oracle Database in order to store and give access to meta-data about CDF data files

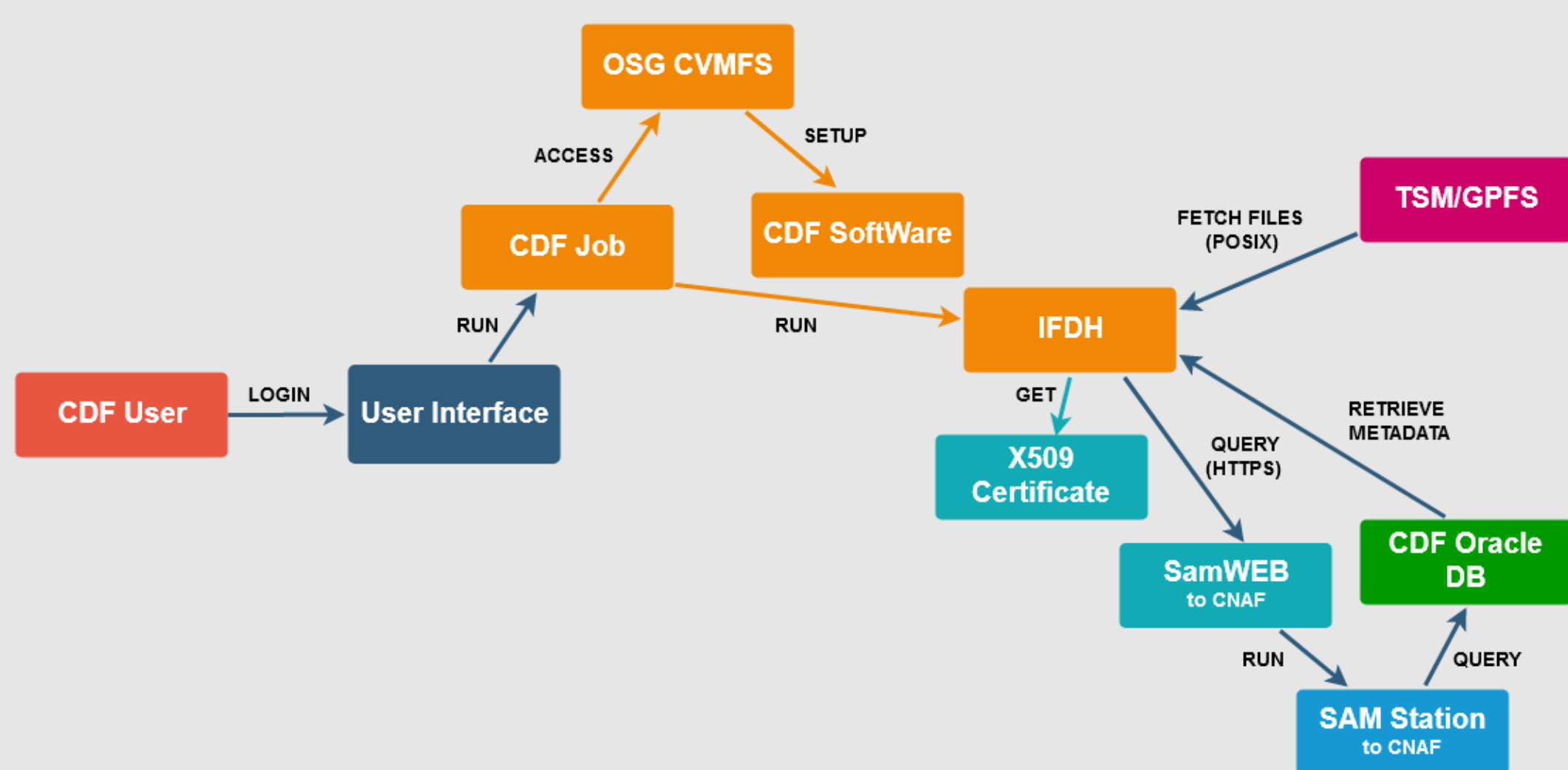


CDF framework at CNAF

FNAL is decommissioning data access framework. CDF Oracle Database has been moved from FNAL to CNAF. CNAF deployed SamWEB and SAM station services and built a framework to access data stored at CNAF or FNAL.

To access data stored at CNAF (most relevant raw and Montecarlo)

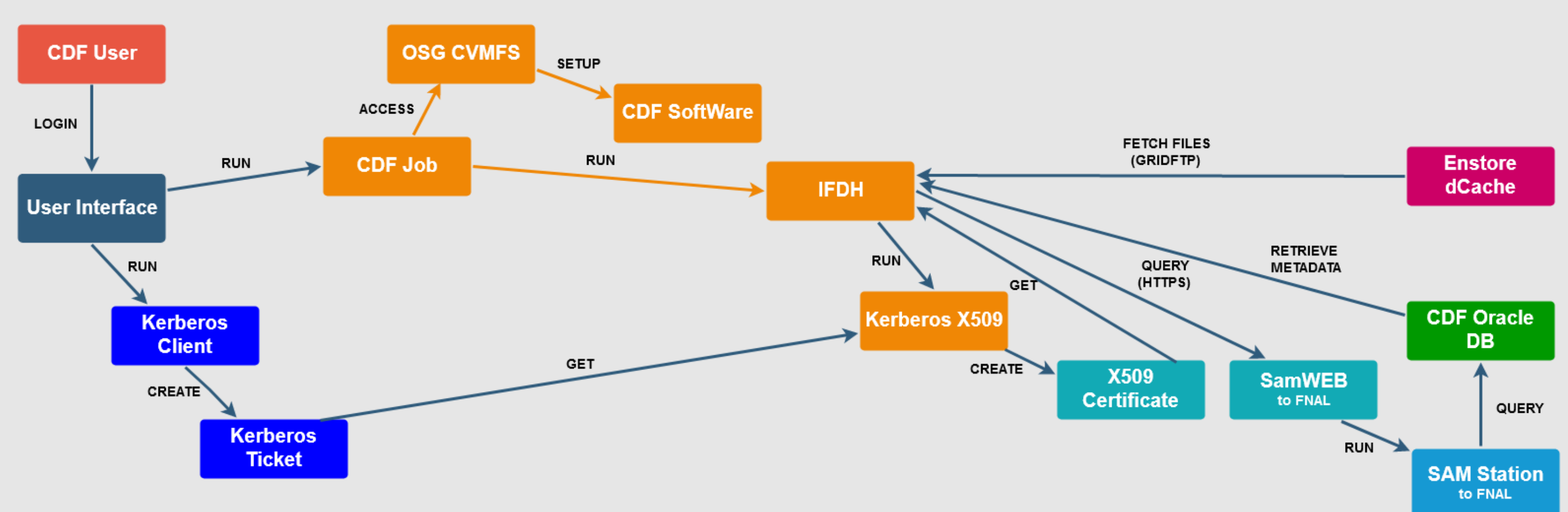
- ✓ The user submits a job that will setup software from Open Science Grid CVMFS
- ✓ IFDH (Intensity Frontier Data Handling) client interacts with SamWEB server (that need a x509 certificate), interfacing with the SAM station, which sends a query to the Oracle Database and gets the proper file locations
- ✓ IFDH copies files from CNAF tape system, through TSM/GPFS service



Flow chart showing CDF files delivery procedure from CNAF tape system

To access data stored at FNAL (all datasets not present at CNAF)

- ✓ FNAL storage system requires an x509 CILogon certificate based on a Kerberos ticket and generated by the user with kx509 command
- ✓ Dedicated SamWEB and SAM station are deployed to retrieve file location at FNAL
- ✓ Data are staged from FNAL tape system, through Enstore/dCache service



Flow chart showing CDF files delivery procedure from FNAL tape system

Tests on framework and data check

- ✓ Functionality tests on CDF framework at CNAF
 - ✓ Some interactive jobs as tests from CDF users properly registered at CNAF and provided with access capability
 - ✓ Collaboration with FNAL Storage Services Administration to solve configuration issues
 - ✓ Tests were successful for data coming from CNAF or FNAL storage systems
- ✓ Data availability and consistency check
 - ✓ All data stored on CNAF tapes have been copied on disk and checksum has been calculated
 - ✓ The installation of the CDF Oracle Database and the SAM station(s) at CNAF allowed to compare the checksum of the files stored at CNAF with the original ones stored in the Database, so that a consistency verification could be applied
 - ✓ All non-consistent files have now been successfully retransferred to CNAF from FNAL Enstore/dCache

Future work

- ✓ Integration of the newly configured HTCondor batch system for CNAF Tier 1
 - ✓ CDF users at FNAL are provided with JobSub, a custom framework developed at Fermilab to submit Condor jobs to remote Condor pools
 - ✓ JobSub manages job submission through a GlideinWMS located at UCSD (University of California San Diego)
 - ✓ Jobs will be executed at CNAF via HTCondor Computing Elements
- ✓ CNAF will provide a high-level service in order to run jobs deciding which SamWEB server to interact with, depending on the actual location of required files (CNAF or FNAL)

